THEME

A Unity and diversity

3 Organisms

From single-celled organisms to coral reefs to trees, life on Earth shows a remarkable degree of variation. For centuries, physical characteristics have been used to name organisms and to put similar organisms into categories. More recently, thanks to DNA sequencing, we can use the genetic code of an organism (its genome) to help show how closely it is related to other species.

HL Cladistics is a way of classifying life forms based on evolutionary relationships. Cladograms are diagrams similar to a family tree but made with species rather than with parents and offspring, and they allow us to study which species are most closely related by examining recent common ancestors.

A3.1 Diversity of organisms

Guiding Questions

What is a species?

What patterns are seen in the diversity of genomes within and between species?

Although there are at least two dozen definitions for the concept of species in biology, we will examine two: the morphological definition that has been used for hundreds of years, and the biological species concept definition, which has only existed in the past few decades. The first looks at what physical features organisms have, while the second considers whether or not individuals can breed to produce fertile offspring. Each definition has its strengths and weaknesses. No single definition can encompass all living organisms as well as extinct species, because such an astoundingly large diversity exists among the various forms of life on Earth.

When DNA sequences of organisms are compared, it is possible to see that, between individuals of the same species, there are remarkably few differences compared to the differences between individuals belonging to two different species. A single-celled organism with no specialized tissue is likely to have a much smaller quantity of DNA than a multicellular organism with hundreds of different specialized tissues.

A3.1.1 – Variation between organisms

A3.1.1 – Variation between organisms as a defining feature of life
Students should understand that no two individuals are identical in all their traits. The patterns of variation are complex and are the basis for naming and classifying organisms.

If you have pigeons where you live, you might think that they all look the same. But ask pigeon experts and they will tell you that the level of diversity and variation among pigeons is equivalent to the level of diversity and variation in humans. Animal breeders such as pigeon fanciers recognize each individual in the population they are raising, just as you would recognize your dog in a group of similar dogs. No two individuals in a population share all the same traits. Even identical twins have slight differences.

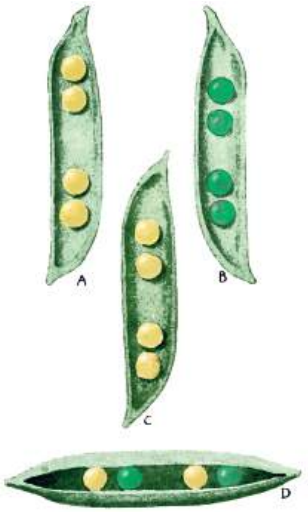
Observing the differences between individuals within one species and observing the differences between one species and another is a daunting task, especially when we consider that there are millions of species on Earth to observe, from invisible microbes to mighty redwood trees over 100 m tall.

How can we classify organisms? There are countless possible ways; a few examples are listed below.

- By feeding habits: it makes its own food/it is a carnivore or herbivore.
- By habitat: land-dwelling/aquatic.
- By movement: sessile (stuck in one place)/free moving.
- By daily activity: nocturnal/diurnal.
- By risk: harmless/venomous.
- By anatomy: plant/animal/vertebrate/invertebrate.

We generally start by categorizing organisms based on **morphology** (the physical appearance of an organism). Is the organism made of a single cell without a nucleus, or does it have a nucleus? If it has a nucleus, is it single-celled or multicellular? Think of these categories as boxes into which the organisms are placed. Each category is called a **taxon** (plural taxa). The biggest taxa are very broad and encompass many species, but as the defining features used become more and more detailed and specific, smaller and smaller boxes are used, containing fewer and fewer species per taxa, until we arrive at a single species. The largest taxon is a "domain" and it contains all the more specific taxa, from "kingdom" down to "species".

Table 1 illustrates the identification of two species from very different kingdoms: one species is an animal, humans, and the other is a plant, garden peas. The science and skill of categorizing life is called **taxonomy** and specialists who do it are called **taxonomists**.



The garden pea (*Pisum sativum*) is the plant Gregor Mendel studied.

Taxa	Human	Garden pea
Kingdom	Animalia	Plantae
Phylum	Chordata	Angiospermophyta
Class	Mammalia	Dicotyledoneae
Order	Primate	Rosales
Family	Hominidae	Papilionaceae
Genus	*Homo*	*Pisum*
Species	*sapiens*	*sativum*

A3.1 Table 1 The classification of two species

How do species exemplify both continuous and discontinuous patterns of variation?

To help remember the order of the taxa, a mnemonic (memory trick) is helpful. Make a sentence using the first letters of each level, such as "King Philip Came Over For Good Soup". The human brain is very poorly adapted for remembering lists of words but very highly adapted for remembering stories. Transforming lists into stories is a good example of a mnemonic.

The variations in characteristics for sorting species into their designated taxon might be obvious (plants have leaves and roots, whereas humans have limbs and a head), but can sometimes be very subtle. Two species of frog might look identical on the outside but can be distinguished by different mating calls. In such a case, the patterns of variation in morphology are not sufficient for classification.

When variation can be placed into distinct categories (type A blood versus type B, for example), we say it is **discontinuous**. When variation has a wide range of possibilities (how tall a tree can grow, for example), we say it is **continuous**. Sometimes we impose categories such as eye colour as if it is an example of discontinuous variation when, in fact, a hundred people who have blue eyes will show a certain amount of continuous variation, from deep blue to very light blue.

A3.1.2 – Species as groups of organisms

A3.1.2 – Species as groups of organisms with shared traits
This is the original morphological concept of the species as used by Linnaeus.

Carolus Linnaeus, an 18th century professor of medicine and botany in Sweden, had difficulty identifying the plants he found on his travels because different botanists used different systems for naming them. This made it difficult to categorize the organisms. Linnaeus then had a remarkable idea: what if we take all the known living organisms, put them into categories, and give them a name using a uniform system? Not just plants, but animals, too. By creating the names using Latin or Greek, no matter what anyone calls the organism in their native language (such as Swedish), it will always have a universally known name.

Linnaeus based the classification system, as well as the names, on the physical features of the organisms. This **morphological classification**, first published in his book *Systema Naturæ* in 1735, was used by generations of botanists and zoologists, and the naming system he created is still used today. Thousands of organisms still carry the scientific name that Linnaeus gave them over two-and-a-half centuries ago, such as the Asian elephant, which he named *Elephas maximus* in 1758.

A3.1.3 – The binomial naming system

A3.1.3 – Binomial system for naming organisms
Students should know that the first part of the name is the genus, the second part of the name is the species. Species in the same genus have similar traits. The genus name is given an initial capital letter but the species name is lowercase.

You have a scientific name based on your species: *Homo sapiens*. This system of naming organisms using two names is called **binomial nomenclature**. "Bi" means two, "nomial" means name and "nomenclature" refers to a system used to name things.

Myrmecophaga tridactyla is a name that literally means "eater of ants" plus "with three fingers". This name refers to the giant anteater of Central and South America. In fact, the animal really has five fingers, but they are hard to see because the animal walks on its front knuckles.

▲ The giant anteater (*Myrmecophaga tridactyla*).

i In the early days of classification, all known organisms were classified into only two kingdoms: plants and animals. With the invention of the microscope in the mid-1600s, many new creatures were discovered that were nothing like plants or animals. In effect, the microscope revealed that there is an entire world of invisible organisms living throughout the world's ecosystems.

The first name in the binomial nomenclature system is always capitalized and it refers to the **genus**; the second name always begins with a small letter and refers to the **species**. Both are always written in italics when typed, or underlined when written by hand. Organisms in the same genus will have a higher number of similar characteristics compared to organisms in a different genus.

There are three main objectives and associated rules to using binomial nomenclature:

1. each organism has a unique name that cannot be confused with another organism
2. the names can be universally understood, no matter what nationality or culture is using the name
3. there is some stability in the system, so that people cannot change the names of organisms without valid reasons.

Examples of binomial nomenclature

Sometimes scientific names for organisms are relatively easy to decipher because they contain their common names:

- *Amoeba amazonas*
- *Equus zebra*
- *Gekko gecko* (this lizard gets its name from the sounds it makes)
- *Gorilla gorilla*
- *Paramecium caudatum* (caudate means having a tail).

Sometimes, it is more difficult to guess their common name:

- *Apis mellifera* (honeybee, although you might have guessed this if you know that beekeeping is also called apiculture)
- *Aptenodytes patagonicus* (king penguin, although you can probably guess where it lives from its species name)
- *Loxodonta cyclotis* (African forest elephant)
- *Malus domestica* (apple tree).

Homo sapiens

Genus ***species***

The rules about writing binomial nomenclature names are that:

- the genus name is capitalized but the species name is not
- both are written in italics when typed, or underlined when handwritten.

SKILLS

In taxonomy, there are two opposing philosophies concerning what to do when an organism does not fit easily into existing categories: (1) broaden the definition of an existing category to include the new organism; or (2) invent a new category or subcategory. Specialists who take the first approach are referred to as **lumpers**, while those who take the second approach are referred to as **splitters**.

Scientists naming organisms sometimes have a sense of humour. Here are some examples.

- *Agra schwarzeneggeri* Erwin, 2002. This Costa Rican ground beetle was named after Arnold Schwarzenegger because of the insect's large biceps.
- *Dracula vampira* Luer, 1978. This orchid in Ecuador got its name from the fact that the petals on the flower look like a bat's wings

Challenge yourself

1. Look up the following to find out what their scientific names are:
   - your favourite animal
   - your favourite fruit or vegetable
   - your favourite flower, tree or house plant.

A3.1.4 – Biological species

A3.1.4 – Biological species concept
According to the biological species concept, a species is a group of organisms that can breed and produce fertile offspring. Include possible challenges associated with this definition of a species and that competing species definitions exist.

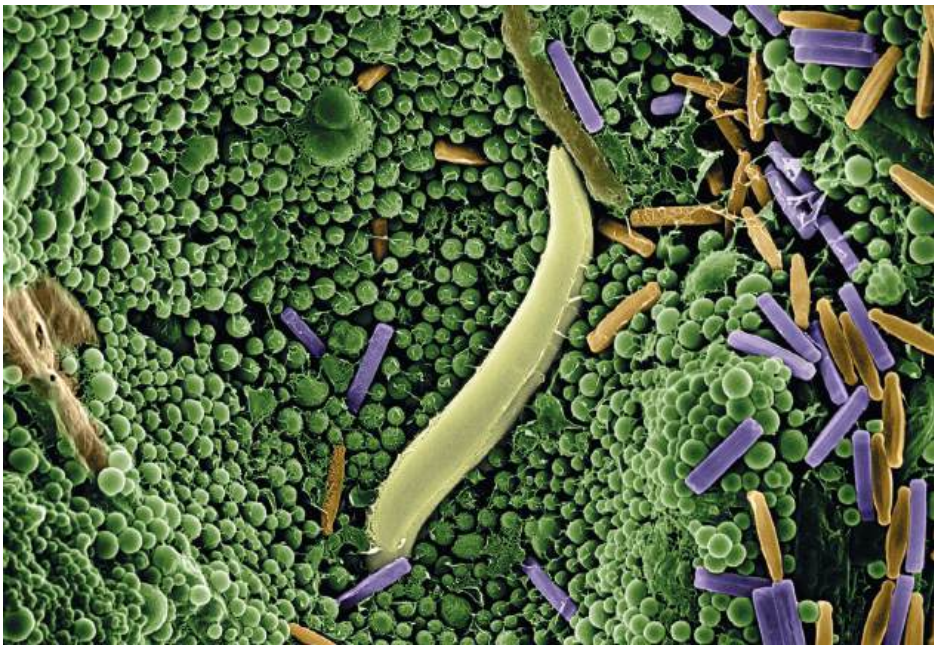
Another definition of a species that is now often preferred over Linnaeus' morphological definition is the **biological species concept**. This was proposed by Ernst Mayr in 1942. Using this definition, in order to be classified as the same species, individuals must be able to breed together and produce fertile offspring. All modern dogs, *Canis familiaris*, can interbreed to produce fertile offspring, so they are considered to be one species.

Not every biologist is happy with this definition, however. How can this definition apply to organisms that reproduce asexually and therefore do not breed? Hybrids produced from parents of closely related but separate species are usually infertile, but not always. Some species are made up of a mosaic of DNA from multiple species. How should they be classified? Should they receive multiple species names if they are composed of more than one? How can we apply the concept to extinct species such as velociraptors when we cannot know from skeletons whether members of a population could interbreed?

All domestic dogs are of the same species.

Depending on which expert you ask, there are dozens of definitions of the word "species". We have discussed two so far: the morphological definition used in the 18th century, and a more recent definition, the biological species concept, involving the ability to breed and produce fertile offspring. But other characteristics can also be taken into account when deciding on what counts as a species, such as the following.

- The ecological niche of an organism. Because microbes are single-celled, it is challenging to use just morphology to determine what species they belong to. Where they live and what they eat can help classify microbes into different species.
- Genetics. When a sequence of DNA found in a sample of soil from a forest does not match any known sample, it suggests that it is from a species that has not been catalogued yet.
- The types of molecules an organism can produce. This is also useful when classifying microscopic organisms that do not have easily observable features, unlike birds and primates, for example. It is common to find microbes that produce carbon dioxide, but some can make methane or hydrogen gas.
- For extinct species, their lineage. If we find a fossil of an extinct snail that has a shell similar to a modern species, we can use the similarities to assign it a species name based on its position on the same part of the evolutionary tree as the existing species.



Microscopic soil organisms can be challenging to identify because morphology is insufficient as a criterion to differentiate species.

Nature of Science

To some extent, the debate about what a species really is becomes just as philosophical as biological. "Is all we are doing simply naming things?" "Do the categories we use actually exist in reality or just in our minds?" "Is the difficulty of agreeing on a definition a fault of the limitations of language?" "Is it possible to use the same term (species) for organisms that exist today and to express how their populations evolved over time?" These questions are currently being debated by biologists and, because the variety of life is so diverse, it is difficult to find a consensus.

A3.1.5 – Distinguishing between populations and species

A3.1.5 – Difficulties distinguishing between populations and species due to divergence of non-interbreeding populations during speciation
Students should understand that speciation is the splitting of one species into two or more. It usually happens gradually rather than by a single act, with populations becoming more and more different in their traits. It can therefore be an arbitrary decision whether two populations are regarded as the same or different species.

Speciation, as explored in more detail in Chapter A4.1, is the process by which a population is separated into two groups that can no longer reproduce together. One part of the population evolves one way and the other, living with different selection pressures and producing different sets of mutations, evolves in a different way. The two populations become different enough over time that they can no longer interbreed to produce fertile offspring. As a result, a new species has branched off from the previous one, resulting in two species that have a common ancestor.

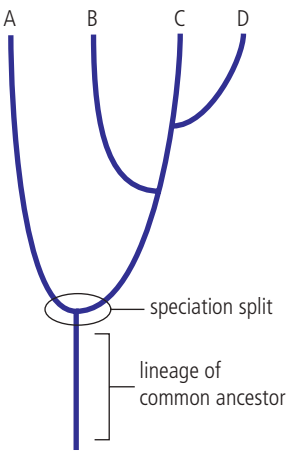
Lake Victoria in East Africa is, geologically speaking, a young lake, being only about 400,000 years old. Any fish species that live there have arrived since then. African cichlid fishes, of which there are over 200 species in the lake, all appear to have evolved from a single species introduced about 200,000 years ago. Each one has evolved in its own niche and as a result split off from the others. Some specialize in eating algae, some eat plankton and others eat snails. But each split would have taken many generations and, during those generations, the population that started to explore the new source of food would have continued to interbreed with some success with the original population. As the two populations became more different from each other, the success rates of interbreeding would have diminished until it was no longer possible. It is difficult for specialists to decide when the speciation occurred. When a cut-off point is chosen, it has an arbitrary and subjective aspect to it.

The last woolly mammoth became extinct thousands of years ago. It appeared to share many similar characteristics with today's Asian elephants (*Elephas maximus*), which is why it was originally classified in 1799 in the same genus, as *Elephas primigenius*. Because of the gap in time, it is difficult to apply the biological species concept to decide whether or not the two populations are one and the same species, because there are no living mammoths to test the hypothesis by breeding them with elephants. The mammoth's scientific name has since been changed to *Mammuthus primigenius*, without knowing for sure whether they could breed together or not, so it is a relatively arbitrary decision from the point of view of the biological species concept.

What might cause a species to persist or go extinct?

The woolly mammoth went extinct thousands of years ago. We cannot test whether it was able to breed with modern elephants or not.

Figure 1 shows a common ancestor giving rise to four species. The first speciation event shown happened earlier in time, then the split that generated species B occurred, and, finally, D split from C. Although this type of diagram helps illustrate the sequence of events, it gives the impression that the splits occurred suddenly, which is not always the case.



A3.1 Figure 1 Species A, B, C and D evolved from a common ancestor. Three speciation splits led to the generation of these species, the first of which is circled.

A3.1.6 – Diversity in chromosome numbers

A3.1.6 – Diversity in chromosome numbers of plant and animal species
Students should know in general that diversity exists. As an example, students should know that humans have 46 chromosomes and chimpanzees have 48. Students are not required to know other specific chromosome numbers but should appreciate that diploid cells have an even number of chromosomes.

Diploid and haploid cells

The term **diploid** is used to describe a nucleus that has chromosomes organized into homologous pairs. Most cells in the human body are diploid cells, and in such cells the nucleus contains a set of 23 chromosomes from the mother and 23 from the father. There is a category of cells that only contain 23 chromosomes in total: the sex cells, also called **gametes**. Because the chromosomes in sperm and egg cells do not come in pairs, but rather only have a single chromosome from each pair, they are said to be **haploid**. The adult form of animal cells is rarely haploid, but there are exceptions, for example adult male bee, wasp and ant cells are haploid. Generally speaking, the vast majority of cells in sexually reproducing organisms are diploid, and only the gametes are haploid.

Note in Table 2 that diploid cells always have an even number of chromosomes. This is logical because one chromosome in each pair comes from one parent and the other from the other parent.

The variable *n* represents the **haploid number**, and it refers to the number of sets of chromosomes that a nucleus can have. For a human egg cell, $n = 23$. When an egg cell is fertilized by a sperm cell (a sperm is also haploid and therefore contains 23 chromosomes), a **zygote** is formed and the two haploid nuclei fuse together, matching up their chromosomes into pairs. Hence humans generally have a total of 23 + 23 = 46 chromosomes. This means that in humans, $2n = 46$, so diploid cells in humans have 23 pairs of chromosomes making a total of 46 chromosomes. Compare this number with some of the other species in Table 2.

A3.1 Table 2 A comparison of types of cells and chromosome numbers

	Types of cells and chromosome numbers	
Species	**Haploid = *n***	**Diploid = 2*n***
Human (*Homo sapiens*)	23	46
Chimpanzee (*Pan troglodytes*)	24	48
Domestic dog (*Canis familiaris*)	39	78
Rice (*Oryza sativa*)	12	24
Roundworm (*Parascaris aquonum*)	1	2

The number of chromosomes is a characteristic of a species

As you can see from Table 2, the number of chromosomes for humans (46) is very different to the number of chromosomes for the roundworm. One of the best-studied worms in genetics laboratories is *Caenorhabditis elegans*, whose genome was first sequenced in 1998. It has six chromosomes, meaning its diploid number, $2n$, is 6, and therefore its haploid number, n, is 3. It would be expected that all the cells in *C. elegans* would have six chromosomes, and, likewise, that all cells in humans would have 46. Although this is true for most cells, we have already seen the exception of haploid cells (n). Note as well that some cells do not contain a nucleus and have no chromosomes, such as red blood cells.

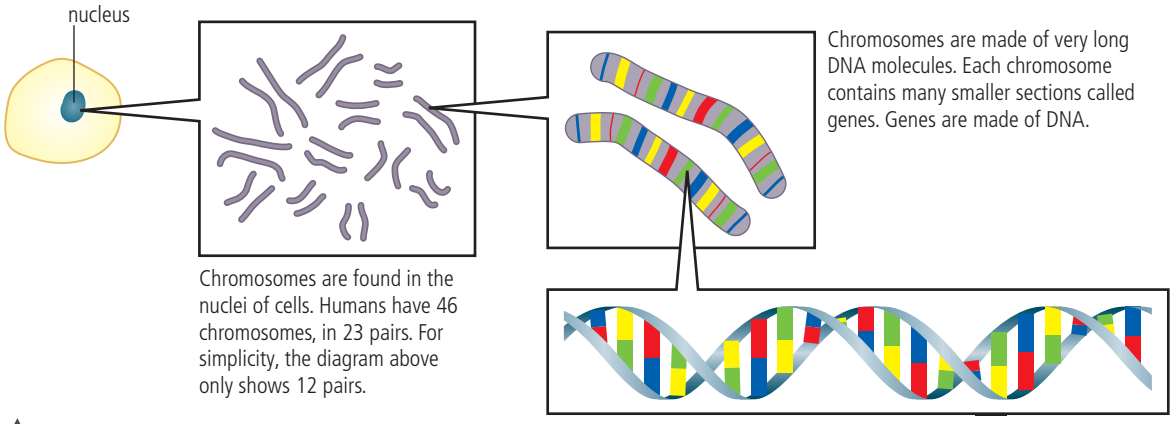
A3.1.7 – Karyotypes

A3.1.7 – Karyotyping and karyograms

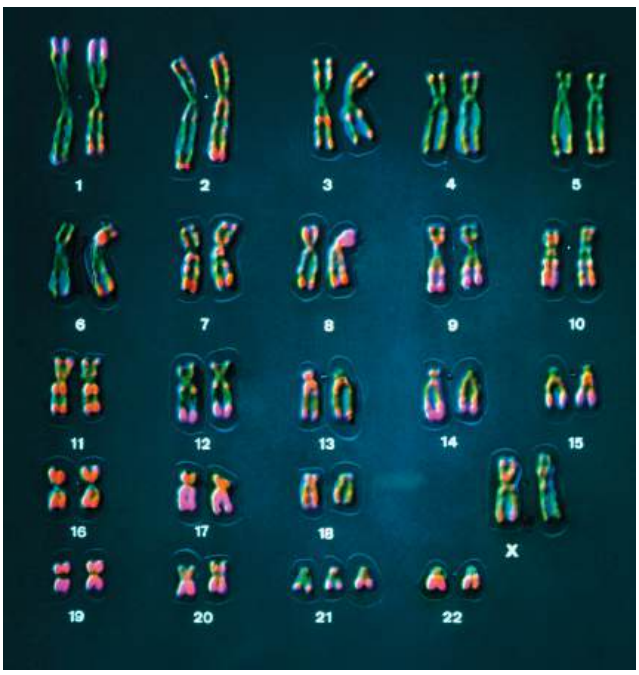
Application of skills: Students should be able to classify chromosomes by banding patterns, length and centromere position. Students should evaluate the evidence for the hypothesis that chromosome 2 in humans arose from the fusion of chromosomes 12 and 13 with a shared primate ancestor.

NOS: Students should be able to distinguish between testable hypotheses such as the origin of chromosome 2 and non-testable statements.

A **karyogram** is a representation of the chromosomes found in a cell arranged according to a standard format, as in the example in Figure 2. The chromosomes are placed in order according to their size and shape. The shape depends mainly on the position of the **centromere**. A karyogram is used to show a person's **karyotype**, which is the specific number and appearance of the chromosomes in their cells.



▲ Zooming into a cell reveals where DNA is found.



◀ **A3.1 Figure 2** This is a karyogram showing all 23 pairs of chromosomes. What can we learn about the individual's karyotype from this figure? This karyogram was prepared using false-colour imagery.

SKILLS

You can use online tools to prepare your own karyogram by arranging chromosomes by size, banding patterns and the position of the centromere. The website Learn.Genetics from the University of Utah has an activity called "Make a karyotype", for example. Once you have made a karyogram, you can learn certain details about the person. Use the karyogram in Figure 2 to determine whether the individual is a male or a female. How do you know? Does the individual's karyotype include any anomalies? If so, describe what you see. For more about the consequences of extra or missing chromosomes, see Chapter D2.1.

How is a karyogram image obtained? Once the cells of an organism have been collected and grown in culture, a karyogram is made following the steps below.

1. The cells are stained and prepared on a glass slide, to see their chromosomes under a light microscope.
2. Photomicrograph images are obtained of the chromosomes during a specific phase of cell division called the mitotic metaphase (see Chapter D2.1).
3. The images are cut out and separated, a process that can be done using a print out and scissors or on a computer.
4. The images of each pair of chromosomes are placed in order by size and the position of their centromeres. Generally speaking, the chromosomes are arranged in order by decreasing length. The exception is in the 23rd pair of chromosomes, which can contain one or two X chromosomes, which are considerably larger than the chromosomes in the 22nd pair (see the chromosome pair marked X in Figure 2). In addition, the coloured bands that show up in the image can be used to identify which chromosome it is. For example, chromosomes 3 and 4 in the image show very different banding patterns.

The evolution of human chromosome 2

Modern humans have 46 chromosomes. Other human species that no longer exist but whose preserved fossil DNA we can study, such as Neanderthals and Denisovans, also had only 46 chromosomes. Gorillas and chimpanzees are the species most closely related to humans. Our last common ancestor with gorillas existed about 9 million years ago and the speciation split with chimpanzees occurred about 6 million years ago. However, when we prepare a karyogram of the contents of their nuclei, both gorillas and chimpanzees have 48 chromosomes instead of 46. If we shared a common ancestor with them, what happened to our chromosome number?

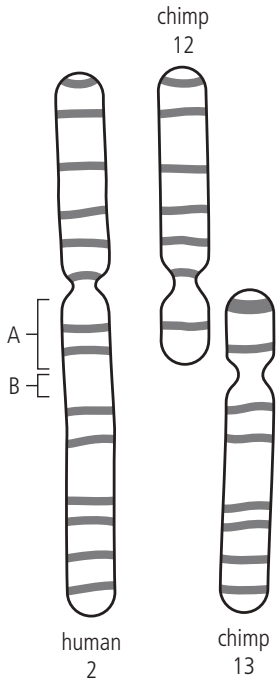
Two possible hypotheses can be formulated:

1. a complete chromosome disappeared
2. two chromosomes from an earlier common ancestor fused to become a single chromosome.

It is unlikely that an entire chromosome was deleted and disappeared, because removing hundreds of genes in that way would cause a major threat to the viability of the species. To test the second hypothesis, we can look for evidence, and can start by examining the two characteristics that help identify a chromosome: its shape (position of the centromere) and its banding patterns. One shape a chromosome can have is the "X" shape, with the centromere close to the centre. This is called a **metacentric** chromosome. Chromosomes can also have an **acrocentric** shape, meaning the centromere is at one end, making one arm of the chromosome much shorter and the other much longer. All primates have both types.

One hypothesis is that chromosome 2 in humans arose from the fusion of chromosomes 12 and 13 in a shared ancestor. In an article from *Molecular Cytogenetics* by Paweł Stankiewicz in 2016, human chromosome 2 was compared to chimpanzee

chromosomes 12 and 13. In terms of shape, these two acrocentric non-human chromosomes, when placed end to end, have a similar length to the human chromosome, although some parts overlap. The position of the centromere in human chromosome 2 lines up with the chimpanzee chromosome 12 but not with chromosome 13. This latter piece of evidence refutes the hypothesis. However, in the zone marked B on the human chromosome in Figure 3, we find the type of DNA we usually encounter in the centromere, known as **satellite DNA**, which consists of short repeating sequences of DNA. This zone corresponds to the position of the centromere in the non-human chromosome 13, giving credibility to the hypothesis. In terms of banding patterns, the long arm of chimpanzee chromosome 12 matches that of the short arm of human chromosome 2, and the long arm of chimpanzee chromosome 13 matches the banding patterns of the long arm of human chromosome 2.



A3.1 Figure 3 A comparison of human chromosome 2 with chimpanzee chromosomes 12 and 13.

Besides shape and banding patterns, other evidence to support the idea of fusion is the presence of telomeric DNA in the centre of human chromosome 2. The **telomeres** are caps at the tips of chromosomes that contain repeating sequences of DNA and provide protection, the same way that bumpers protect cars and aglets protect the ends of shoelaces. Such repeating telomeric DNA is not supposed to be in the centre of chromosomes, only at the tips. And yet, at position A in the human chromosome 2 shown in Figure 3, telomeric DNA is present at the position where the two chromosomes would have fused.

It is very important to understand that this evidence does not say we descended from chimpanzees. The fusion of the chromosomes would have happened after the speciation split of a common ancestor that led to the evolution of chimpanzees on one branch of the tree of life and the evolution of humans on another branch.

When asked to evaluate evidence for a claim, scientists and students need to express their opinion of whether or not the evidence is sufficient to convincingly confirm the claim. Some questions to consider asking are:

- Is the quantity of evidence sufficient to accept the claim?
- Has the method for collecting evidence been repeated and tested by other scientists, and have they found similar evidence?
- Is the method being used a reliable method?
- Are any counterclaims or refuting evidence enough to doubt the claim?
- Is there a mechanism to explain the cause, or is what we are seeing just a coincidence?

Nature of Science

Some claims are testable and others are not. The hominid fossil nicknamed Lucy, discovered in Ethiopia in 1974, is complete enough to test and confirm claims such as (1) she was a female, (2) she was not a modern human but rather an australopithecine, (3) she could walk on two legs and (4) she lived about 3.2 million years ago. There might be some debate about the details, but the challenges can also be tested. Can you think of any claims about her that would not be testable? For example: "Lucy had a great sense of humour." "Lucy had a recurring dream where she encountered a wildcat." "Lucy spoke three languages." Current tools in science have no way of testing these claims. Statements like these are speculation. What about these: "Lucy had very little meat in her diet." "Australopithecines such as Lucy had strong spiritual beliefs." Are they testable claims?

Some claims about the fossil called Lucy are testable and others are not.

A3.1.8 – Unity and diversity of genomes

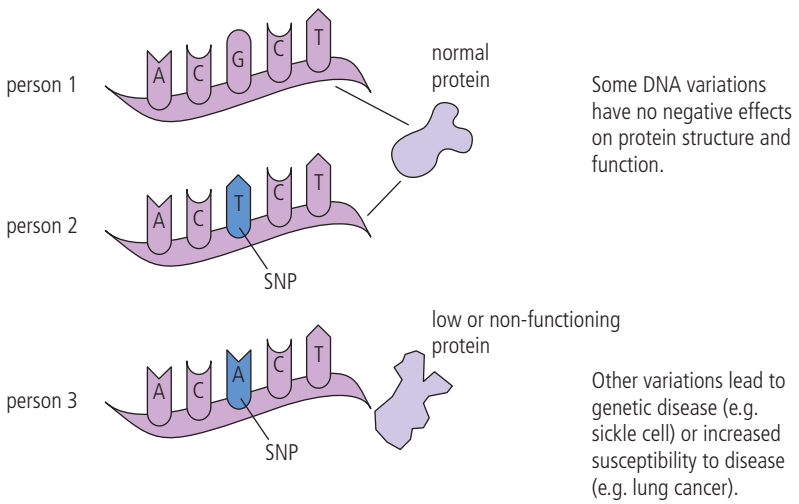
A3.1.8 – Unity and diversity of genomes within species
Students should understand that the genome is all the genetic information of an organism. Organisms in the same species share most of their genome but variations such as single-nucleotide polymorphisms give some diversity.

It seems counterintuitive, but it is possible to find lots of evidence to support the claim "we are all the same", and it is also possible to find lots of evidence to support the claim "we are all different". From a genetics point of view, humans share many more similarities than differences with each other, especially compared to another species.

If a chimpanzee was walking down your street, you would recognize right away that it was a non-human primate. And yet, the genetic difference between us and chimpanzees is only about 4%. That is a much bigger difference, however, than between you and other humans, which is estimated to be 0.1% to 0.6%. Why does *Homo sapiens* display so many similarities within its global population? Our unity arises

largely from the fact that all humans share the same genes. We do not all have the same versions of each of the genes (called **alleles**, see Chapter D3.2); some of us have type B blood and some have type O, for example. But we all possess the genes that determine the ABO blood type.

Where do we find these small but crucial differences between humans? The estimated 3 million to 20 million **base pairs** (e.g. A–T or G–C) of our DNA sequence that can reveal the differences are found scattered all over our chromosomes. Where most people have a T (thymine) nucleotide, for example, a small portion of humans might have a G (guanine) instead at that position. Such variations can start out as mutations (see Chapter D1.3) but are then passed down from generation to generation. Such a variation involving only one base is called a **single nucleotide polymorphism** or SNP (see Figure 4). It is estimated that about every 100 to 300 bases in a human's genetic code contains an SNP. Geneticists interested in the human genome have identified millions of SNPs, and they can be used to help determine ancestry or risk of genetic diseases.



A3.1 Figure 4 Person 1 has a gene that expresses a normal protein. Person 2 has a T (thymine) nucleotide instead of a G (guanine) in the SNP, but also expresses a normal protein. Person 3, however, has an SNP that causes the protein to not form correctly.

Only about 5% of SNPs are functional, meaning they actually produce a difference in a person's body. Most are neutral, meaning that they will not affect a person's **phenotype** (the physical expression of a gene, such as blood type or colour vision, see Chapter D3.2).

The Human Genome Project

In 1990, an international cooperative venture called the Human Genome Project set out to sequence the complete human **genome**. Because the genome of an organism is a catalogue of all the bases it possesses, the Human Genome Project hoped to determine the order of all the bases A, T, C and G in human DNA. As there were approximately 3,200,000,000 to find, it took over a decade. In 2003, the Project announced that it had succeeded in achieving its goal. Now, scientists are working on deciphering which sequences represent genes and which genes do what. The human genome can be thought of as a map that can be used to show the position of any gene on any one of the 23 pairs of chromosomes.

i In the 1997 science fiction film *GATTACA*, one of the main characters brings a sample of cells to a walk-up window at an establishment that provides anonymous genome services. Within seconds, she gets a full printout and analysis of the genome she is interested in. How far are we from being able to do this today? What ethical implications are there to such a service? Are there laws protecting your genome?

Thanks to modern communication technologies, it is possible for scientists working all over the world to collaborate and contribute to a scientific endeavour such as sequencing the genome of plants that help feed the world. Rice is one example: biologists from 10 countries contributed to sequencing the first rice genome.

The current estimate is that humans have approximately 22,000 genes, and, thanks to advances in technology, the sequencing of a person's genome can be done in hours instead of years.

TOK

Many companies offer genome sequencing for private citizens willing to pay the price. Some of the products reveal ancient family origins and risk factors for some health problems, such as the chances of developing certain types of cancer or heart disease. Would you want to know if there is a chance that your life could be suddenly shortened by the presence or absence of a certain gene? Would you tell your family and friends? Would you want your parents to have such a test? Should people tell their employer or each other about any health-related issues revealed by a genomic analysis? Or, in contrast, is this a private, personal thing that no one else needs to know about? How accurate and reliable are these analyses? Should we believe everything they say? Does all knowledge impose ethical obligations on those who know it?

A3.1.9 – Eukaryote genomes

A3.1.9 – Diversity of eukaryote genomes
Genomes vary in overall size, which is determined by the total amount of DNA. Genomes also vary in base sequence. Variation between species is much larger than variation within a species.

No humans have genes for characteristics such as bioluminescence (glowing in the dark), which many deep-sea organisms do. Although we see some diversity among humans, we do not see such huge ranges of diversity in the human population as wings for flight, gills to breathe underwater, echolocation organs for seeing without light, chloroplasts for photosynthesis, and so on. There is more unity within the human species (comparing any two humans) than diversity compared to other species (comparing humans to non-humans).

Humans are a diverse global population but there are remarkably few differences between any two humans compared to differences between humans and other species.

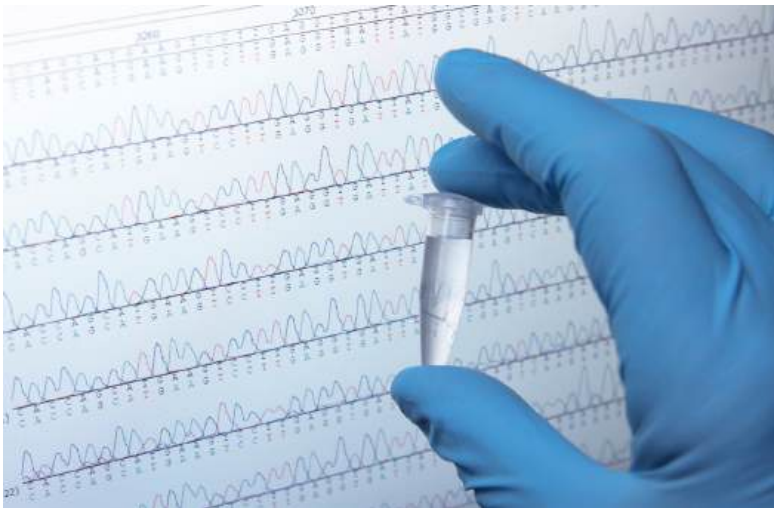
One major difference between genomes is their size: the quantity of DNA they have in their nuclei. As we will see in Section A3.1.10, some eukaryotic genomes only have a few thousand genes while others can have tens of thousands of genes. This means that one eukaryote will possess genes that another will not have at all. A fish does not

need to have genes to produce pollen, and a rose bush does not need genes for making fins to swim. Even with closely related species that have undergone a relatively recent speciation split, they have been evolving separately to the point where the genes are now different enough that they cannot interbreed anymore.

Such differences can be seen in the sequences of base pairs in each genome. Sequencing technology along with databases and computer programs for searching and comparing large data sets have allowed biologists to compare the genomes of organisms from all over the world.

Bioinformatics is a research field that uses both computer science and information technology to help us understand biological processes. Bioinformatics has grown exponentially in recent years. The most data-rich area of bioinformatics is genomics. Genome data is now available in public databases such as The National Center for Biotechnology Information (NCBI). Genetic information can also be explored using the following databases:

- Swiss-Prot, a database of protein sequences
- Ensembl, a database and browser of genomic information about humans and other vertebrates
- GenBank, a National Institutes of Health genetic sequence database that is an annotated collection of all publicly available DNA sequences.



A micropipette containing a DNA sample can be sequenced and added to a database and shared worldwide thanks to web-based information technology.

Instead of sifting through the entire genome of an organism, one way to compare genetic diversity in eukaryotes is to focus on their **mitochondrial DNA**. All eukaryotes have mitochondria, and the way mitochondrial DNA, present only in the egg, not in the sperm cell, is passed down from mother to offspring, means there is not the shuffling and mixing that we see in chromosomal DNA. It is estimated that, within a species, roughly 1 in 1,000 of the genetic code letters is different between individuals' mitochondrial DNA. These genetic differences are expressed in the amino acid sequence that is coded for by the organism's DNA sequence. To see differences between individuals within a species, or to see differences between species, it is possible to look up the amino acid sequences for a particular gene in a database and match them to see if there are amino acids missing, added or modified. Instead of the DNA bases A, T, C and G being displayed, the letters in the databases correspond to the 20 possible amino acids, such as S for serine, G for glycine, A for alanine and V

for valine. Some amino acids have a letter that is different from their first letter, such as E for glutamic acid, F for phenylalanine and K for lysine. You will not be asked to memorize the 20 amino acid names and their letters, but you do need to understand that, when comparing genetic differences, it is possible to either use the DNA code or the amino acid sequences.

Table 3 shows part of the sequence for a single gene selected from the online UniProt protein database. The chosen gene is one that all eukaryotes have in their DNA: *cyc1*, the gene for cytochrome c, which is a protein needed by mitochondria to perform their essential task of cellular respiration, to convert sugar into usable energy. Of the hundreds of species available in the database, four species of animal were selected and, rather than looking at all the amino acids that the gene codes for, a short sequence of 60 amino acids was selected for comparison. The differences between the first species and the three other species are highlighted in yellow.

A3.1 Table 3 Comparing a short sequence of 60 amino acids from the mitochondrial gene, *cyc1*, for cytochrome c, in four species

Database codes for specific species	Fragment of the sequence of amino acids coded for in the *cyc1* gene
golden-crowned babbler:	
TR|A0A7K9SBC6|A0A7K9SBC6_9PASS	SL--ALALSLGGGPLSAGELELHPPNFPWSHGGPLSALDHASVRRGFQVYRQVCSACHSM
brown-headed cowbird:	
TR|A0A7L3VSC4|A0A7L3VSC4_MOLAT	SLAVALSLSLGGGPVSAGELELHPPGLPWSHGGFLSALDHASVRRGFQVYRQVCSACHSM
green anole:	
TR|H9GCG1|H9GCG1_ANOCA	GLAVALH-----SAVSAGELELHPPSFPWSHSGPLSSLDHSSVRRGYQVYKQVCSACHSM
big-headed turtle:	
TR|A0A4D9DRJ9|A0A4D9DRJ9_9SAUR	GLALALH-----TAVSASDLELHPPSYAWSHNGLLASLDHSSIRRGYQVYKQVCAACHSM

The first organism in Table 3 is a bird, the golden-crowned babbler (*Sterrhoptilus dennistouni*), which lives in the Philippines. The next three organisms in Table 3 are a brown-headed cowbird (*Molothrus ater*), a lizard called a green anole (*Anolis carolinensis*), and a big-headed turtle that lives in Southeast Asia (*Platysternon megacephalum*). If we look at the first amino acid in the sequence for the first species, we see S, for serine. Moving down the second column in Table 3, we see that species 2 also has an S but species 3 and 4 have G for glycine instead. Species 1 does not have any amino acids at positions three and four, while the other three do. Of those three, they all have A for alanine in the third position but not all have V for valine in the fourth.

Not surprisingly, compared to the first bird's sequence, there are more differences in the lizard and in the turtle than there are in the other bird species, because the two bird species are more closely related to each other than they are to lizards and turtles. If we looked at the whole amino acid sequence and not just the fragment of 60 amino acids used for Table 3, we would see that the three species in Table 3 have the following percentage of matches with the golden-crowned babbler: 92.9%, 84% and 76.8%, respectively.

Between any two golden-crowned babblers, we would expect more than 99% of the amino acid sequence to be identical, with only one difference every few hundred amino acids. This illustrates that there is much more diversity between organisms in different species compared to organisms within the same species.

The Human Genome Project has shown that there are only a very small number of DNA bases that make one person different from any other person in the world. This creates a feeling of unity. All humans carry inside them a common genetic heritage.
On the other hand, the Human Genome Project has shown that the small differences that do exist make each person unique in terms of skin colour, facial features and resistance to disease, for example. These differences should be appreciated and celebrated as strengths. Unfortunately, they are often the basis of discrimination and misunderstanding.
Can one group of people be considered genetically superior to another? History has shown that many people think so, yet genetics shows that this is not the case. All human populations, whatever slight differences their genomes may have, deserve equal esteem as human beings.

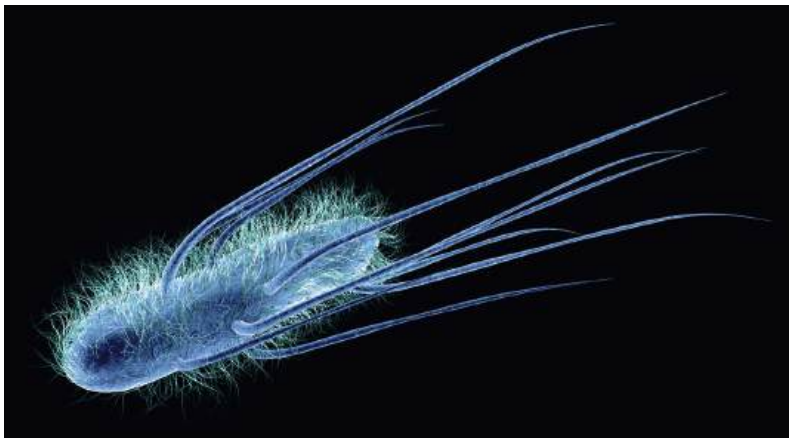
A3.1.10 – Genome sizes

A3.1.10 – Comparison of genome sizes
Application of skills: Students should extract information about genome size for different taxonomic groups from a database to compare genome size to organism complexity.

Using online tools, it is possible to compare the genome of an organism, such as a fruit fly, with other eukaryotes. Table 4 shows data extracted from the NCBI database at the time of writing; because the database is being continually updated, the numbers you find might be different.

A3.1 Table 4 A comparison of genome sizes of various organisms

Species	Genome size in millions of base pairs, Mb
Saccharomyces cerevisiae, baker's yeast	12.1
Drosophila melanogaster, fruit fly	143.7
Mus musculus, house mouse	2,500
Escherichia coli, bacterium	5.12
Homo sapiens, modern human	3,200
Neoceratodus forsteri, Australian lungfish	34,557.6
Plasmodium falciparum, a parasite that causes malaria	22.9
Oryza sativa, rice	420
Caenorhabditis elegans, a nematode worm	100



Escherichia coli, a bacterium that likes to live in your large intestine, has about 5 million letters (base pairs) in its DNA code.

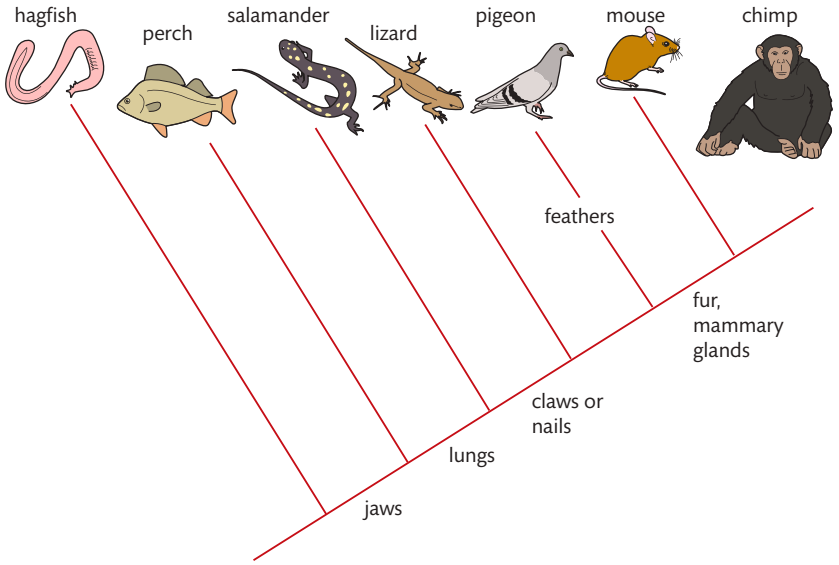
Do you get the impression that the more complex an organism is, the bigger its genome is? For example, we think of humans as being extremely complex and advanced, so when we compare ourselves to the fungus in Table 4, the baker's yeast, we see that our genome size is hundreds of times bigger. But rice has only three times more DNA than the fruit fly. And when we compare our human genome size to the Australian lungfish, it is ten times smaller. Does that mean lungfish are more complex than we are or that we are more complex than yeast? It depends on our definition of complex. Although they may not be capable of doing creative and complex tasks such as sending a spaceship to Mars, both lungfish and yeast can survive in conditions in which humans would die. The examples given and the ones you can find on your own will often give the impression that genome size can indicate complexity, but there are enough exceptions to conclude that it is not a reliable indicator.

A3.1.11 – Whole genome sequencing

A3.1.11 – Current and potential future uses of whole genome sequencing
Include the increasing speed and decreasing costs. For current uses, include research into evolutionary relationships and for potential future uses, include personalized medicine.

Researchers are very excited about genome sequencing because it allows them to identify species and compare them to see evolutionary relationships. They can compare whole genome sequences to see how organisms are related to each other. Such a technique is known as **phylogenetics**. In general, organisms that share similar genomes tend to be more closely related than those that do not.

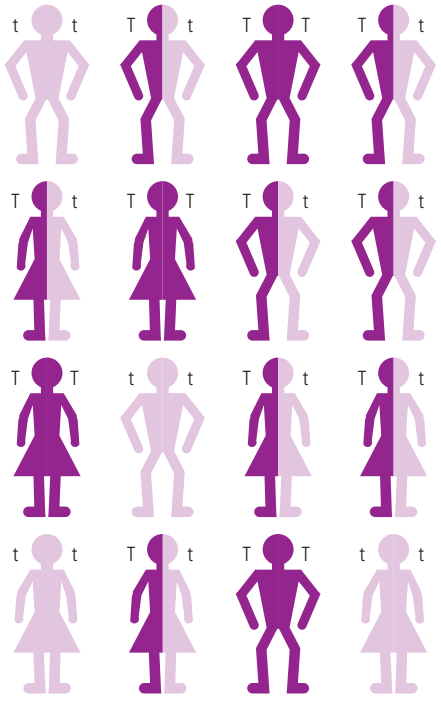
In Figure 5, the mouse is shown to be much more closely related to the chimpanzee than to the salamander. The DNA sequences (or corresponding amino acid sequences) of the mouse and the chimpanzee would show fewer differences between each other than if one of their DNA sequences was compared to the salamander's genome. In humans, it can tell us about our ancestry, and about possible health risks related to the genes we have inherited.



A3.1 Figure 5 A phylogenic tree of vertebrate chordates.

Thanks to **next-generation sequencing techniques**, which use a mix of laboratory hardware, chemical markers and powerful software to increase the speed and decrease the cost of sequencing people's genomes, it is possible for private citizens in some countries to get their genomes sequenced. Other countries have made it illegal to request genome sequencing: laws have been put in place to protect people's privacy. A parent who has put up a child for adoption and does not wish to be identified, for example, might have their identity revealed by this technology even if they do not have their own genome scanned, because a close relative's genome might be sufficient to find the match. In other countries, such services are fully legal and gaining popularity. Several companies in the United States offer genomic testing and provide detailed reports about ancestry and possible health issues related to DNA.

One potential such sequencing holds is the concept of **personalized medicine**, sometimes called precision medicine: information about a person's genetic makeup can be applied to an individual when prescribing treatments. The premise is that, if doctors know a patient's DNA profile, the best adapted treatment can be prescribed. When a doctor prescribes a drug today, the choice of molecule and the dose is based on studies involving people who might not be representative of everyone's genetic makeup. By sequencing the genomes of the participants in drug trials, patterns can be identified that suggest one drug might work better with people who possess a particular genetic sequence, but that for others, another molecule, combination of drugs or different dose would provide better results or perhaps fewer side effects.



A3.1 Figure 6 Knowing that a particular medication produces severe side effects only in people who receive the t version of an identified gene from both parents (*tt*) would allow doctors to know that four people in this group of patients should not be prescribed that medication. All the other patients have received a T from at least one parent (they are either *TT* or *Tt*) and can benefit from the medication without severe side effects.

Personalized medicine is better adapted for diseases that are dynamic, such as cancer, type 2 diabetes or cardiovascular disease, and require different treatments at different stages of the illness. Knowing more about how a patient's genome might cause new proteins to be produced in their cells or trigger certain genes to be turned on or off could lead to breakthroughs in medical treatments. By creating databases of biomarker profiles within a population (such as ***TT***, ***Tt*** or ***tt*** in the example in Figure 6), researchers of personalized medicine hope to provide better diagnoses and more effective treatments with fewer undesirable side effects.

Another advantageous use of the human genome is the production of new medications. This process involves several steps:

- find beneficial molecules that are produced naturally in healthy people
- find out which gene controls the synthesis of a desirable molecule
- copy that gene and use it to instruct synthesis of the molecule in a laboratory
- distribute the beneficial therapeutic protein as a new medical treatment.

This is not science fiction: genetic engineering firms are finding such genes regularly. One current line of research is dealing with genes that control ageing. How much money do you think people would be willing to pay for a molecule that could reverse the effects of ageing and prolong life by several decades?

HL

A3.1.12 – Difficulties with the biological species concept

A3.1.12 – Difficulties applying the biological species concept to asexually reproducing species and to bacteria that have horizontal gene transfer

The biological species concept does not work well with groups of organisms that do not breed sexually or where genes can be transferred from one species to another.

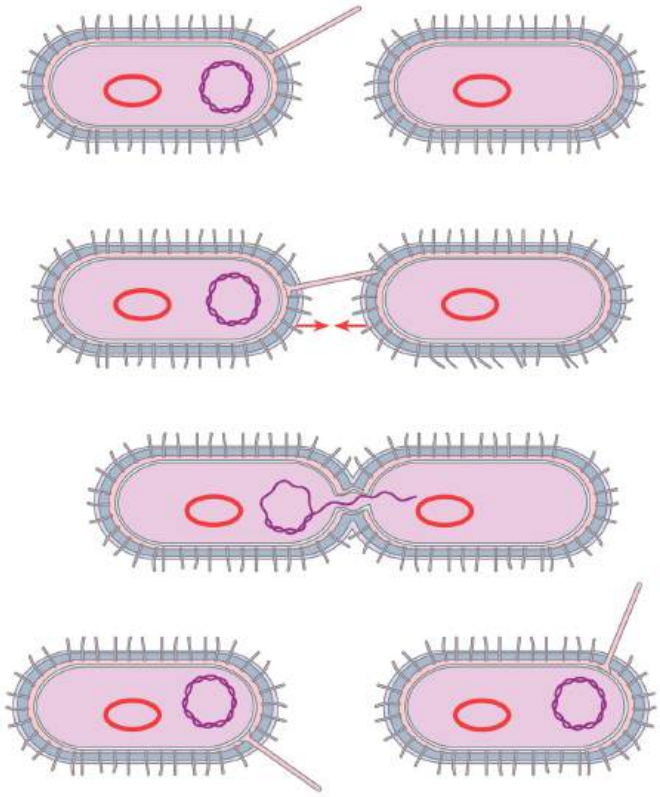
Insects usually reproduce by having males fertilize the eggs of females. The females of certain stick insects in the genus *Phasmatodea*, however, can often produce young without mating with a male. The eggs mature and grow into adult females. This process is called **parthenogenesis**. A similar process happens in plants called **vegetative propagation**, such as when a strawberry plant sends out a runner that takes root near the original plant. Farmers can plant last year's potatoes in their fields to grow new potato plants from them. In such cases, each new plant is an identical copy of the parent plant. This could continue for many generations without the need for sexual reproduction. These generations are continuing to produce offspring but they are not doing it by breeding, so they pose a challenge to the biological definition of what a species is.

Strawberry plants can clone themselves by sending out runners that become new plants.

The examples mentioned can either breed by mixing gametes or produce clones by making copies of themselves, but some organisms can only reproduce asexually. Bacteria reproduce **asexually** using **binary fission** rather than by breeding. There is no such thing as mother/father or male/female, and no gametes are produced. Bacterial cells grow larger, make a copy of their genetic material, and split into two daughter cells that are identical to the original parent cell.

The idea of passing on genes to the next generation is one of the cornerstones of biology. The direction of such a transfer of genes is vertical, from one generation down to the next. But in addition to passing down their genes, bacteria can undergo **horizontal gene transfer**. Whereas other organisms normally only receive genetic material once in their life, when the male and female sex cells that formed them first meet and fuse together, bacteria can exchange genetic material within their lifetime, rather than just at the start. If a bacterium in a population has a mutation that could be useful to another member of the population, the two cells can attach to each other and

exchange sequences of DNA. Remarkably, this gene transfer can be done even if the host bacterium is *not* of the same species.



▲ The bacterium on the left, the donor cell, is passing genetic information to the bacterium on the right, the host cell, in a process called **plasmid transfer**.

One of the major assumptions of the concept of a species is that all members of a species have a common lineage and come from a series of common ancestors. This is the basis of the **tree of life** concept. The idea of mixing lineages by trading genes complicates things hugely. When a bacterium has a mix of genes from its own species that it received from previous generations, mixed with genes from a donor species during its lifetime, it poses a challenge to the idea of sharing a common ancestry with the other members of its species. If horizontal gene transfer happened several times in previous generations and again during its lifespan, the genetic material inside the bacterium would be a mosaic of genes from various sources.

Another challenge to the definition of species is the gene transfer that can occur from bacteria to archaea, from viruses to eukaryotes, and bacteria to eukaryotes. We can accept the idea that we all have inside our human cells organelles that were once prokaryotes (e.g. our mitochondria, see Chapter A2.2), and we know that we contain virus DNA, so who are we, really? Can we say that we are pure human, or are we a mosaic of genes from both human and non-human sources?

When sequencing and matching genes, sometimes a sequence of DNA is found that has more in common with another species than the one it is found in. Such genes, known as **xenologs** or **jumping genes**, travel in plasmids from one bacterium to another. Sometimes we find the same identical gene in several very different species

TOK

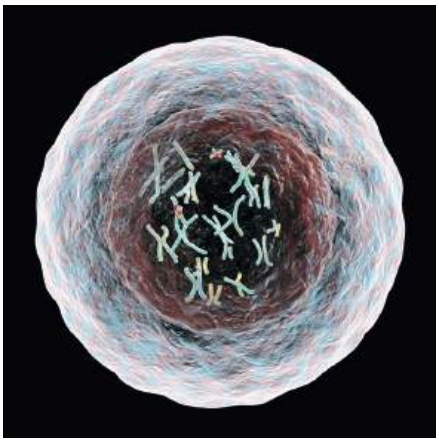
The tree of life concept is a good example of a well-accepted way of seeing how life evolved over time. From a central point representing the earliest forms of life, biologists believe the tree then branches out in all directions, and each split that creates a new branch is a speciation. Genetic information flows from one generation to the next and, once a split has occurred, there is no going back. The concept of horizontal gene transfer is an example of a paradigm shift, challenging the long-held beliefs of biologists. Evidence accumulated over the last 100 years challenges the original concept of the tree of life, and instead suggests that the tree is, in fact, more like a web. Genetic information can jump from branch to branch, causing an interweaving of the branches into a mesh. Organisms not only share common ancestors but also have unexpected ancestors from completely different branches. How does the way that we organize or classify knowledge affect what we know?

of bacteria. Usually when we see similar DNA sequences, we think that the organism shares a common ancestry but, in this case, the host species are from different branches of the tree of life and do not share the same lineages, such as yeast cells (a fungus) containing bacterial DNA. Such examples challenge the concept of species because it questions the idea that organisms classified in the same genus or species have a common ancestor.

A3.1.13 – Chromosome number as a shared trait

A3.1.13 – Chromosome number as a shared trait within a species
Cross-breeding between closely related species is unlikely to produce fertile offspring if parent chromosome numbers are different.

All domestic dogs belong to the same species, *Canis familiaris*, and dogs have a chromosome number of 78. The pineapple (*Ananas comosus*), on the other hand, has 50 chromosomes. The fruit fly (*Drosophila melanogaster*) has a chromosome number of 8. The red king crab (*Paralithodes camtschaticus*) has 208. You do not need to memorize these numbers, but you do need to know how the number of chromosomes a species possesses is a characteristic that all members of that particular species share. Of course, there are exceptions, which are explored in Chapter D2.1, but when we state the chromosome number of a species, such as 46 for humans, we are referring to the typical diploid number expected in the cells of that species.



Although there are exceptions, generally speaking, the number of chromosomes found in an organism's cells should be identical between all members of the same species.

A female horse and a male donkey can mate and produce a mule. However, mules cannot usually mate to make more mules. Because the offspring (the mules) are not fertile, no new species has been created. Instead, a mule is called an **interspecific hybrid**. Hybrids face several challenges to continue as a population. For one thing, the vast majority of animal and plant hybrids are infertile. Even if one generation of hybrids is produced, a second generation is highly unlikely. This presents a genetic barrier between species.

Look at the chromosome number (in parentheses) of a mule and its parents:

female horse (64) + male donkey (62) = mule (63)

Mongolian wild horses (the Przewalski's horse, *Equus przewalskii*) and domesticated horses (*Equus caballus*), which have 66 and 64 chromosomes, respectively, have been known to produce hybrids as well.

What happens when the chromosome number of the parents is different? A mule born with a chromosome number of 63, which is neither that of the mother species nor that of the father species, makes it challenging to successfully mate within the populations of either. The atypical chromosome number makes it difficult for homologous pairs of chromosomes to match up during meiosis (discussed in Section D2.1.9), and thus production of gametes can be difficult. Mules therefore cannot pass on their genes to a subsequent generation, and therefore are not considered a new species. However, in rare cases, interspecific hybrids have been observed to produce fertile offspring.

A3.1.14 – Dichotomous keys

A3.1.14 – Engagement with local plant or animal species to develop a dichotomous key
Application of skills: Students should engage with local plant or animal species to develop a dichotomous key.

When biologists encounter an organism they need to identify, they can use a **dichotomous key** to establish which taxa it belongs to. If you have ever played a guessing game in which the rule is that you can only ask "yes" or "no" questions, then you already know how a dichotomous key works.

Here are the basic principles of how to use a dichotomous key. You can try it out with the example in Figure 7.

1. Look at the first section of the key, which has a pair of sentences, (a) and (b), describing characteristics.
2. Next, look at the organism to see if the particular characteristic described in the first line (a) is present in the organism.
3. If the answer is yes, then go to the end of that line and find the number of the next pair of statements to look at, follow the number given and continue until the end. If the end of the line contains a name, it is the taxon for the organism.
4. If the answer is no, then go to the second statement just below it, (b), and that one should be true; go to the end of that line and find the number of the next pair of statements to look at. Follow the number given and continue until the end.

The idea is to keep going until you get to a name instead of a number: if you have answered each question correctly, that will be the name of the taxon your organism belongs to.

Dichotomous key

1	a)	No differentiated tissues, no symmetry or identifiable organs	Porifera (sponges)
	b)	Presence of differentiated tissues and organs	go to 2
2	a)	Stinging cells present, can show radial symmetry	Cnidaria (e.g. sea jellies)
	b)	No stinging cells	go to 3
3	a)	Has two-way digestive tract and bilateral symmetry	Platyhelminthes (flatworms)
	b)	Has a one-way digestive tract	go to 4…

etc…

A3.1 Figure 7 Example of the beginning of a dichotomous key to identify animal taxa.

SKILLS

Develop your own dichotomous key for local plant or animal species. Full details of how to carry out this activity with a worksheet are available in the eBook.

A3.1.15 – DNA barcoding

A3.1.15 – Identification of species from environmental DNA in a habitat using barcodes
Using barcodes and environmental DNA allows the biodiversity of habitats to be investigated rapidly.

When we go to a store to buy a T-shirt or a bottle of orange juice, the barcode on the label is scanned and converted to a number. That number is used to identify the item by matching it against a database of items that includes their prices. Using a similar idea, genetic sequences obtained from organisms can be given a number (a **barcode identification number** or BIN) that is matched against a database of sequences that are known to belong to previously identified and named organisms. A **DNA barcode** is a short sequence of DNA (several hundred base pairs) inside an organism's cells that can be used to quickly identify the species.

To barcode a specimen means to sequence its genetic material and match a specific sequence to a known sequence stored in a genetic library. Thanks to the technology of DNA sequencing, millions of these barcodes have been added to libraries. A category that is commonly used to identify animals is mitochondrial DNA. For prokaryotes, sequences found in ribosomes (ribosomal RNA rather than DNA) are used for the barcodes instead.

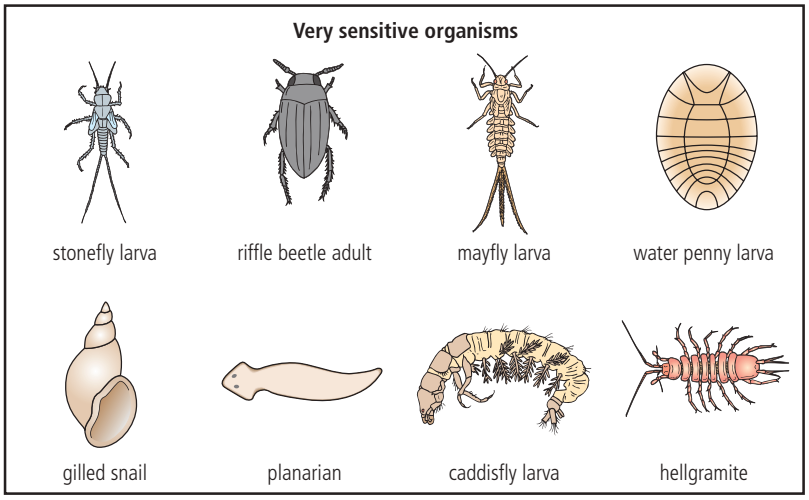
In order to be usable by scientists and researchers everywhere, the data about these barcodes needs to be stored in a place that is accessible, such as the Barcode of Life Data System, or BOLD, developed in Canada. If there is a very strong match (99% or more), then we have a high level of confidence that the correct species name has been found for an organism. If the match is less strong, we have less confidence that it is the same species, and other techniques should be used to confirm the true species.

Such databases can be used to rapidly identify the various species present in an ecosystem. Water samples can be taken from lakes, rivers, estuaries and oceans, and soil samples can be taken from fields and forests. The DNA extracted from the water or soil is sequenced and barcodes are isolated for analysis. Such DNA collected from the environment rather than from an organism is called **environmental DNA** or **eDNA**. It is present in the environment because organisms release dead cells, produce faeces or die and start to decay. Think of a wild animal shedding hair, a tree losing a leaf or a fish releasing waste into a river. DNA can be found in any cells shed by an organism. After separating out the different DNA in a sample, it can be amplified using a technique called **polymerase chain reaction (PCR)** (see Chapter D1.3) and then sequenced.

Water or soil samples collected in an ecosystem can be sequenced in a laboratory to identify the organisms that inhabit it.

Specialists studying a zone near a polluted area often want to know if the **biodiversity** of the zone is affected. Biodiversity can be measured by identifying and counting the number of species present. Although dichotomous keys and experienced expert eyes can identify organisms reasonably quickly down to the family or genus level, it can be difficult to identify to the species level with a high degree of accuracy because of very subtle differences in appearance or in differences that might not be noticeable at the time of observation. For animals such as insects, the larval stage often looks very different from the adult stage. For plants, if an unidentified plant is not flowering or a tree has lost its leaves, identification using morphology can be very challenging if not impossible because these features are often crucial to their proper identification. To identify organisms morphologically, they have to be observed directly or captured, which can be time consuming as many animals are elusive. Identifying organisms by physical features is also significantly more difficult for microbes, which cannot be seen unless you use a microscope or can culture them in a laboratory. With DNA barcodes and fast sequencing technology, samples of DNA can be sequenced and matched to identify species with a high level of confidence in a matter of hours.

Certain species can be used as **bioindicators** or **indicator species**: these organisms are so sensitive to certain types of pollution that their presence in an ecosystem indicates a lack of pollution. Conversely, their sudden disappearance from an ecosystem would suggest the appearance of a source of pollution. Caddisfly larva live underwater in streams and can be used as bioindicators for water health. A large caddisfly population is reassuring, but a small or disappearing caddisfly population is a signal to investigators that they should start looking for potential contamination upstream. DNA barcoding of water samples can indicate the presence of one or more species of caddisfly.



Examples of indicator species that are very sensitive to pollution and whose presence in an ecosystem can reassure us that the habitat is healthy.

To determine whether biodiversity is increasing or decreasing, a baseline is needed to compare to later: we need to know how many species are present now so that we can see if the number goes up or down later. In most parts of the world, terrestrial, aquatic or marine ecosystems do not have such baselines, so we cannot know how much biodiversity is changing. Barcoding of environmental DNA can help rectify this. Look through the TOK prompts. Do any apply to this example? What are the implications of having, or not having, knowledge?

TOK

To perform an ecological survey, one species at a time is identified and counted, and often, for example with insects, birds or fish, this requires capturing them, which can disrupt the population. With eDNA metabarcoding, a single sample can be sequenced for dozens or hundreds of species without having to capture individual organisms. This is more time efficient for the specialists carrying out the ecological surveys. Over months and years, if the number of species in an ecosystem decreases, we know that the biodiversity of that habitat is unhealthy or declining. In contrast, if the number of species is stable or increases over time, we know that biodiversity is stable or improving, suggesting that the ecosystem is healthy and flourishing.

There are some disadvantages to using environmental DNA. Firstly, it only gives an indication of the presence or absence of a species, not the population size. Secondly, the DNA does not indicate if it is from a living organism or a dead one. Thirdly, certain chemical incompatibilities exist with the processing of soil samples because substances in the soil can interfere with the sequencing process, giving rise to inaccurate results.

HL end

Testing for genetic material shed by organisms can be applied in unexpected situations. In the second year of the COVID pandemic, in addition to testing people for viruses, many municipalities were testing wastewater for the presence of SARS-CoV-2. In New York City, some sequences that were identified were from strains of the virus that had never shown up in samples from patients. This suggested that the problem was more complex and widespread than previously understood. Towns and cities that used this technique could target certain neighbourhoods for additional testing, prevention and medical resources.

Guiding Question revisited

What is a species?

In this chapter we have learned that:

- there is no single definition of the term "species" because the sheer variety of currently living species and extinct species is so enormous and complex
- using morphology works up to a point, but this methodology is poorly adapted for microbes or for species that are visually very similar
- the biological species concept works most of the time but is does not work for single-celled organisms that do not breed, or for organisms that are only found in the fossil record.

Guiding Question revisited

What patterns are seen in the diversity of genomes within and between species?

In this chapter we have discussed how:

- there is some diversity in genomes of individuals of the same species
- there is much more diversity when two different species are compared, especially if they were separated in a speciation event that occurred long ago.

Exercises

Q1. The system of giving a scientific or Latin name to organisms such as *Canis familiaris* is used worldwide. State the name of this system and identify the person who perfected and popularized it.

Q2. Distinguish between the morphological definition of species and the biological species concept.

Q3. Explain the features of chromosomes that are taken into consideration when making a karyogram.

Q4. Distinguish between haploid and diploid cells.

Q5. A karyogram can be used to determine if an unborn baby will be a girl or a boy. Explain how a karyogram is analysed to do this.

Q6. Outline the evidence for a fusion of ancestral chromosomes to become human chromosome 2.

Q7. Outline the advantages of personalized medicine using genomes.

Q8. HL Discuss reasons for and against using environmental DNA and barcoding for ecological surveys of biodiversity.